DOCUMENT RESUME

ED 308 202 TM 013 409

AUTHOR Nandakumar, Ratna
TITLE Traditional Dimensionality vs. Essential Dimensionality.
PUB DATE Mar 89
NOTE 14p.; Paper presented at the Annual Meeting of the American Educational Research Association (San Francisco, March 27-31, 1989).
PUB TYPE Reports - Evaluative/Feasibility (142) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Definitions; Educational Assessment; Latent Trait Theory; Mathematical Models; *Monte Carlo Methods; Psychometrics; *Simulation; *Statistical Analysis; Testing Problems
IDENTIFIERS *Dimensional Analysis; Item Parameters; *Unidimensionality (Tests)

ABSTRACT

The theoretical differences between the traditional definition of dimensionality and the more recently defined notion of essential dimensionality are presented. Monte Carlo simulations are used to demonstrate the utility of W. F. Stout's procedure to assess the essential unidimensionality of the latent space underlying a set of terms. The traditional definition of dimensionality makes no distinction between major and minor dimensions. It is desirable to count only dominant dimensions in psychometric assessment of dimensionality of a latent space. Stout (1987, 1988) has provided a definition known as essential dimensionality and has developed a statistical test to assess the essential unidimensionality of a set of items. Monte Carlo simulation studies with two examinee sizes, n=750 and n=2,000, were generated with item parameters resembling those of real tests. Both simulations exhibited good performance in assessing essential unidimensionality. Since applications of item response theory techniques are becoming increasingly popular in most educational assessment, procedures to assess unidimensionality accurately will greatly facilitate the use of item response theory. Four tables provide data from the simulations. (SLD)

TRADITIONAL DIMENSIONALITY VS. ESSENTIAL DIMENSIONALITY

Ratna Nandakumar

University of Delaware

Paper presented at the annual meeting of the American Educational Research Association, San Francisco, March, 1989

Abstract

This paper points out the theoretical differences between the traditional dimensionality and essential dimensionality, empirically demonstrating through Monte Carlo simulations the utility of Stout's procedure to assess, statistically, essential unidimensionality underlying a set of items.

Presently, most item response theory (IRT) based research assumes one–, two–, or three–parameter logistic or normal ogive models. The two most critical assumptions underlying these models are "unidimensionality" and "local independence." In other words, these models can be applied to estimate an individual's latent ability provided the test is measuring only one ability or dimension: and for a given ability level, examinees' responses to different items are independent. The intent of this paper is to present the differences between the traditional definition of dimensionality and the more recently defined notion of essential dimensionality, and to empirically demonstrate, through Monte Carlo simulations, the utility of Stout's procedure to assess essential unidimensionality of the latent space underlying a set of items.

Definition of Dimensionality

Let U_i denote the i–th item, and $\underline{U}_N \equiv (U_1, U_2, \ldots U_N)$, the test. Observed values of item and test will be denoted by u_i and $\underline{u}_N \equiv (u_1, u_2, \ldots u_N)$ respectively. Let $U_i = 1$ denote a correct response and $U_i = 0$ denote an incorrect response to item i for a randomly chosen examinee. The latent random vector is denoted by $\underline{\Theta}$ and the particular values it takes is denoted by $\underline{\theta}$. Let $P_i(\underline{\theta})$ denote the probability that a randomly chosen examinee with ability $\underline{\theta}$ will get the i–th item correct. Let $X = \{U_i, i \geq 1\}$ denote an item pool.

According to the classical notion of dimensionality (Lord & Novick, 1968) if k traits influence n items in a test, that is, each of the k traits influence examinees' performance on at least one item in the test, then k is the dimensionality of the latent space. These k traits represent k psychological dimensions and define a k–dimensional complete latent space.

The above definition, although mathematically precise, makes no distinction between major and minor dimensions. It has been argued that items are inherently multiple determined and it is not uncommon to find more than one ability influencing

relatively few items or is unique to individual items in a test. These insignificant attributes should not be included in psychometric assessment of dimensionality (Humphreys, 1984). More over, there could be other factors such as test anxiety, motivation, ability to work quickly etc., in addition to the main dominant dimension that influence the test performance of an examinee (Hambleton & Swaminathan, 1985). Hence it is desirable to count only dominant dimensions in psychometric assessment of dimensionality of a latent space.

Example: Consider the construction of a paragraph comprehension test of n items. Let n=5r, denote the total number of items in a test where 5 denotes the number of paragraphs and r denotes the number of questions following each paragraph. Let the first paragraph be about British history, the second about the Second World War, the third about the Sahara desert, the fourth about classical music, and the fifth paragraph about art history.

Clearly in the above example reading is the common ability influencing all the items. In addition each paragraph is influenced by one other ability, namely British history in the first paragraph, events in the Second World War in the second paragraph, and so on. Clearly, according to the traditional definition of dimensionality, the number of dimensions underlying the above set of items is 6, wherein it is desirable to assess the dimensionality as 1, namely reading, which is the dominant dimension.

Although it has been implicitly assumed that only dominant dimensions be considered in assessing the dimensionality of a latent space and several procedures based on factor analytic notion have been developed to assess the dimensionality of a latent space, until recently, there has been no precise definition as to what is meant by dominant dimension. Stout (1987, 1988), for the first time, has provided such a definition known as

essential dimensionality which flows from the definition of *essential independence* as opposed to traditional assumption of local independence. Stout (1987) also has provided a statistical test procedure to assess essential unidimensionality of a set of items. Nandakumar (1987) has further refined Stout's procedure. Stout (in press) argues that traditional assumptions of unidimensionality and local independence can be replaced by respective weaker and psychometrically more appropriate assumptions: essential unidimensionality and essential independence. Junker (1988) has proved mathematical results linking local independence and essential independence, and has also provided procedures for consistently estimating ability under essential independence.

Definition 1 (Stout, 1988). The item pool X is said to be *essentially independent* (EI) with respect to latent variables $\underline{\Theta}$ if X satisfies

$$D_N(\underline{\theta}) \equiv \frac{\sum \sum_{1 \le i < j \le N} \left| Cov(U_i, U_j | \underline{\Theta} = \underline{\theta}) \right|}{\binom{N}{2}} \to 0 \text{ as } N \to \infty.$$

The distinction between LI and EI is that LI requires $Cov(U_i, U_j | \underline{\Theta} = \underline{\theta}) = 0$ for all $\underline{\theta}$, whereas EI requires that the average $Cov(U_i, U_j | \underline{\Theta} = \underline{\theta})$ is small in magnitude (for all $\underline{\theta}$) asymptotically. Hence EI is a weaker assumption than LI.

Definition 2 (Stout, 1988). The *essential dimensionality* (d_E) of an item pool X is the minimal dimensionality necessary to satisfy the assumption of EI. When d_E=1, essential dimensionality is said to hold.

Method

Monte Carlo simulation studies are used to demonstrate the effectiveness of Stout's procedure (Stout, 1987; Nandakumar 1987) to assess essential unidimensionality. Two types of tests, as described below, are considered where they are essentially unidimensional ($d_E=1$) but require more than one single ability in order to correctly answer the items. In both cases a bivariate extension of the unidimensional three parameter logistic model with compensatory abilities (Equation 1) is used to generate item responses with uncorrelated abilities. Two examinee sizes: J=750 and J=2000 are simulated in each case. The item parameters are generated to resemble those of real tests. The means and the standard deviations of simulated items are like those of real tests obtained from different sources in the literature.

$$(1) \qquad P_i(\theta_1, \theta_2) = c_i + \frac{1-c_i}{1 + \exp\{1.7[a_{1i}(\theta_1-b_{1i})+a_{2i}(\theta_2-b_{2i})]\}}$$

Case 1

This case is analogous to paragraph comprehension test. The n items of a test are split into h groups where each group of items is influenced by two abilities. Out of these two abilities, one ability is common to all items of the test (example, reading), and the other ability is unique to each group (example, art history, second world war etc.). In all there are 1+h abilities and each item is influenced by two abilities. The degree of influence of common ability is considered major and the degree of influence of second ability is considered minor. The item discriminating parameters are generated such that their means and variances are a function of the degree to which they influence the trait; and the ratios of means of a_1 and a_2, and the ratio of variances of a_1 and a_2 are the same. For instance, if

the contribution of the major ability is .8 and the contribution of the minor ability is .2. then the means and variances of a_1 and a_2 are a function of the degree of contribution to their respective abilities. The contribution of minor ability is denoted as Cona2. The contribution of major ability is 1–Cona2. Abilities are generated from standard normal distribution.

Case 2

This is different from Case 1 in that there are only two abilities one major and the other minor and the same two abilities influence all items. The item and ability parameters are generated as in the case of Case 1.

Results

The simulation results shown in Table 1. 2. and 3 exhibit good performance of Stout's procedure in assessing essential unidimensionality. Each cell of Table 1 displays rejection rate for the Case 1 with two items per paragraph. per 100 repeated trials of Stout's procedure. on simulated tests that resembles conceptually that of a paragraph comprehension test. According to the asymptotic theory when $d_E=1$. one expects on the average about 5% rejections. Each cell of Table 2 displays rejection rate per 100 repeated trials for the Case 1 with five items per paragraph. In other words each minor ability influences 5 items. Contents of Table 1 and Table 2 indicate that Stout's procedure is able to establish essential unidimensionality. For short tests of 25 or 30 items however, r=5 amounts to one fifth of the test being contaminated. Stout's procedure in such a case assesses it as not essentially unidimensional. Table 3 displays rejection rates for the Case 2. where only two abilities. one major and one minor. influence all items of the test. The results of Table 3 are in accordance with what is desirable. These results can be contrasted with those of a typical two–dimensional case as displayed in Table 4 (Stout. 1987). where a test consists of two

major abilities influencing all items. The results of Table 4 indicate that Stout's procedure maintains very good power when the correlation between abilities is as high as 0.5.

Educational Implications of the Study

Since applications of IRT techniques are becoming increasingly popular in most educational assessment procedures, uncritical use of IRT models can have grave consequences in estimating and interpretations of educational outcomes. Unidimensionality being one of the critical assumptions of IRT modeling, procedures to accurately assess unidimensionality will greatly facilitate the use of IRT in both theoretical as well as applied research.

References

Hambleton. R. K. & Swaminathan. H. (1985) Item Response Theory: Principles and Applications, Kluver–Nyjhoff Publishers, Boston.

Humphreys. L.G. (1984) A theoretical and empirical study of the psychometric assessment of psychological test dimensionality and bias. ONR Research Proposal.

Junker, B. (1988). Some Foundational Results in Ability Estimation, Ph.D. Thesis. Department of Statistics. University of Illinois at Urbana–Champaign

Lord. F. M. & Novick. M. R. (1968) Statistical Theories of Mental Test Scores. Reading Massachusetts: Addison –Wesley.

Nandakumar, R. (1987) Refinement of Stout's procedure for assessing latent trait dimensionality. Doctoral dissertation . University of Illinois. Urbana–Champaign.

Stout. W. F. (1987) A non parametric approach for assessing latent trait unidimensionality. Psychometrika. 52, 589–617.

Stout, W. F. (in press) A new item response theory modeling approach with applications to unidimensionality assessment and ability estimation. Psychometrika.

Table 1

Rejection Rates per 100 Trials for the Case 1: One Major and Several Minor abilities influence test items. $d_E = 1$, Cona2[a] = 0.2, r[b] = 2, $\alpha = .05$

J	TESTS		
	n=26	n=40	n=50
750	1	3	1
2000	9	2	2

Notes:

[a]Cona2 denotes the contribution of minor ability influencing test items.

r[b] denotes the number of items influenced by a particular minor ability

J denotes the number of examinees simulated.

n denotes the number of items in a test.

Table 2

Rejection Rates per 100 Trials for the Case 1: One Major and Several Minor abilities influence test items, d_E = 1, Cona2[a] = 0.2, r[b] = 5, α = .05

J	TESTS			
	n=25	n=30	n=40	n=50
750	11	7	3	1
2000	27	15	3	1

Notes:

[a]Cona2 denotes the contribution of minor ability influencing test items.

[b]r denotes the number of items influenced by a particular minor ability

J denotes the number of examinees simulated.

n denotes the number of items in a test.

Table 3

Rejection Rates per 100 Trials for the Case 2: One Major and One Minor abilities influence test items. $d_E = 1$, Cona2[a] = 0.2, $\alpha = .05$

J	TESTS
	n=50
750	5
2000	3

Notes:

[a]Cona2 denotes the contribution of minor ability influencing test items.

J denotes the number of examinees simulated.

n denotes the number of items in a test.

TABLE 4

Rejection Rates for d = 2, c = 0.2 Logistic Study
Number of Rejections per 100 Trials

Test			SATV		ACTM		ACTE			ASVAB AS		ASVAB AR
M			8	12	7		8		12	5	7	7
$N_1:N_2:N_3$			17:17:16	17:17:16	13:13:14	35:5:0	17:17:16	40:10:0	40:10:0	8:8:9	8:8:9	10:10:10
J	ρ	α										
750	0.5	0.01	44		37	33	34	46				58
		0.05	62		69	58	59	71				76
		0.10	71		83	70	76	76				84
	0.7	0.01	17									
		0.05	36									
		0.10	48									
2000	0.5	0.01	91	94			75	99	93	68	77	
		0.05	96	99			90	99	96	80	94	
		0.10	96	100			91	100	98	88	98	
	0.7	0.01	58	74	52	29	27			36	32	57
		0.05	81	85	74	49	55			53	54	67
		0.10	86	89	86	60	65			62	61	74